AARHUS
UNIVERSITY

CENTER FOR QUANTITATIVE
GENETICS AND GENOMICS

Genomic prediction using a G-matrix weighted with SNP variances from Bayesian mixture model

Guosheng Su, Ole F. Christensen, Luc Janss
and Mogens S. Lund

Background

- Bayesian variable selection models often predict more accurate genomic breeding value (GEBV) than GBLUP
- GBLUP has low computational demand
- One approach to combine the advantages of both models is to use weighted G-matrix (Zhang et al. 2010)
- Various weighting factors for building G-matrix

Objective

- Achieve the benefits of both GBLUP and Bayesian variable selection models by using weighted G-matrix

- Assess weighting factors and methods to build G-matrix.

- Investigate the time-interval to update weights

Statistical models and methods

1. GBLUP: GBLUP model using original G-matrix

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Zg} + \mathbf{e}$$

g is the vector of additive genetic effects, $\mathbf{g} \sim \mathrm{N}(\mathbf{0}, \mathbf{G}\sigma_g^2)$

$\mathbf{G}=\mathbf{MM}'/n_{marker}$ (VanRaden 2008)

$$\mathbf{m}_j = \frac{0-2p_j}{\sqrt{2p_jq_j}},\ \frac{1-2p_j}{\sqrt{2p_jq_j}},\ \frac{2-2p_j}{\sqrt{2p_jq_j}}$$

for genotype A_1A_1, A_1A_2 and A_2A_2, $p=p_{A2}$

2. BMIX: Bayesian mixture model assuming a mixture of 4 distributions for SNP effects (Gao et al. 2013)

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Mq} + \mathbf{e}$$

q is the vector of SNP effects.

$$q_i \sim \pi_1 N(0,\sigma_1^2) + \pi_2 N(0,\sigma_2^2) + \pi_3 N(0,\sigma_3^2) + \pi_4 N(0,\sigma_4^2)$$

$\pi_1=0.889$, $\pi_2=0.1$, $\pi_3=0.01$, $\pi_4=0.001$

$$\sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2$$

3. GBLUP$_{\mathbf{w}}$: GBLUP model using weighted G-matrix to account for heterogeneous variances of SNP effects

$\mathbf{y} = \mathbf{1}\mu + \mathbf{Zg} + \mathbf{e}$

g is the vector of additive genetic effects, $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G_w}\sigma_g^2)$

$\mathbf{G_w} = \mathbf{MTM'}/n_{marker}$

T is a diagonal matrix with t_{jj}= weight on SNP j

Weighting factors

1. Square of individual SNP effect from the BMIX (Wq^2)
2. Posterior variance of individual SNP from the BMIX model (WV_q), calculated as $\Sigma\pi_i\sigma_i^2$

Final weight is standardized so that mean weight = 1

Weighting methods

1. Single-marker weighting: one weight for one marker
2. Group-marker weighting: one common weight to a group of markers

A common weight = mean weight of the makers in the same group

Group sizes: 5, 10, 30, 50, 70, 100, and 150

Data analysis

Marker and Phenotypic Data – Nordic Holsteins

Marker data: 54K chip

Phenotypic data: De-regressed proofs (DRP) of milk, fat, protein, fertility, mastitis

Reference data: About 3950 bulls born before 2005

Test data: About 1150 bulls born in 2005 and later

Data to derive weight

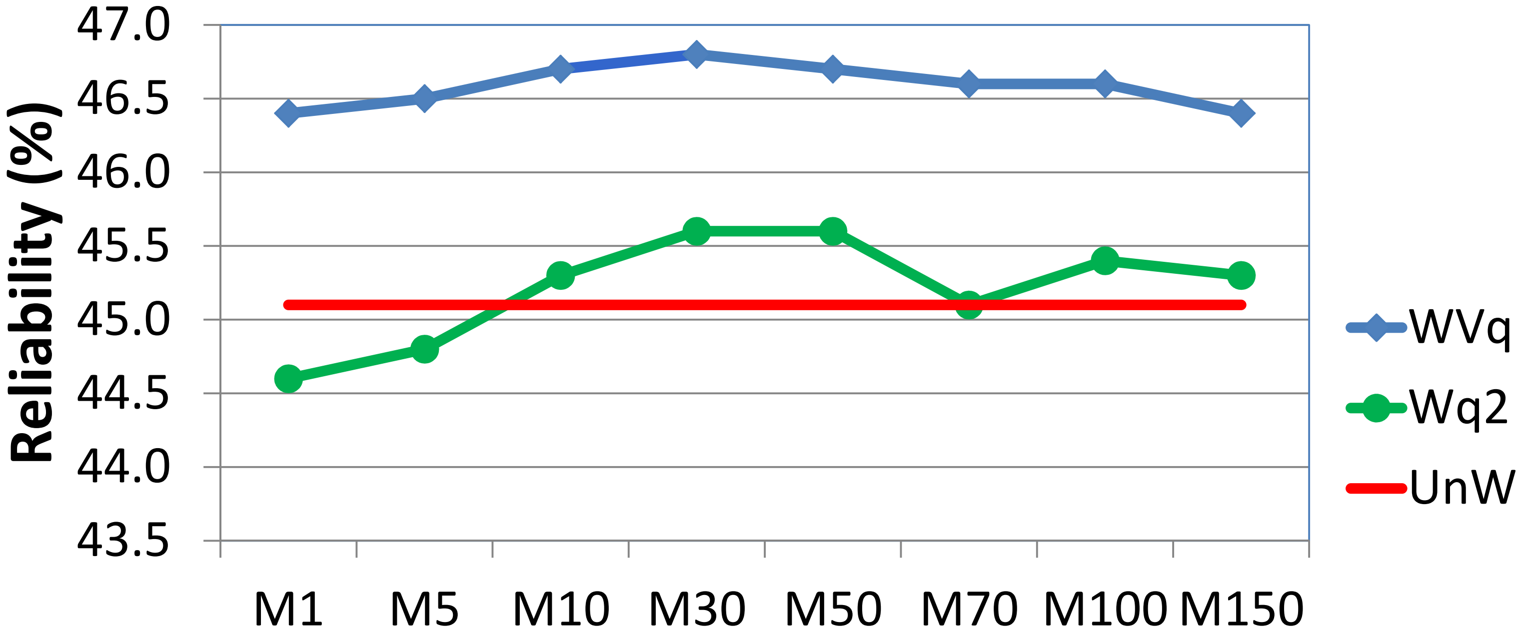
1. **Lag0y**: The same as the reference data (before 2005) used for current prediction
2. **Lag1y**: Bulls born before 2004, i.e., 1 year before current prediction
3. **Lag3y**: Bulls born before 2002, i.e., 3 years before current prediction
4. **Lag5y**: Bulls born before 2000, i.e., 5 years before current prediction

Reliability of GEBV and regression coefficient of DRP on GEBV, obtained from GBLUP and BMIX4

Trait	Reliability (%)		Regression	
	GBLUP	BMIX4	GBLUP	BMIX4
Milk	48.3	51.6	0.87	0.88
Fat	46.8	50.8	0.84	0.83
Protein	46.2	47.8	0.81	0.82
Fertility	44.6	45.1	0.98	0.97
Mastitis	39.5	40.4	0.90	0.90
Mean	**45.1**	**47.1**	**0.88**	**0.88**

BMIX gave 2.0% higher reliability

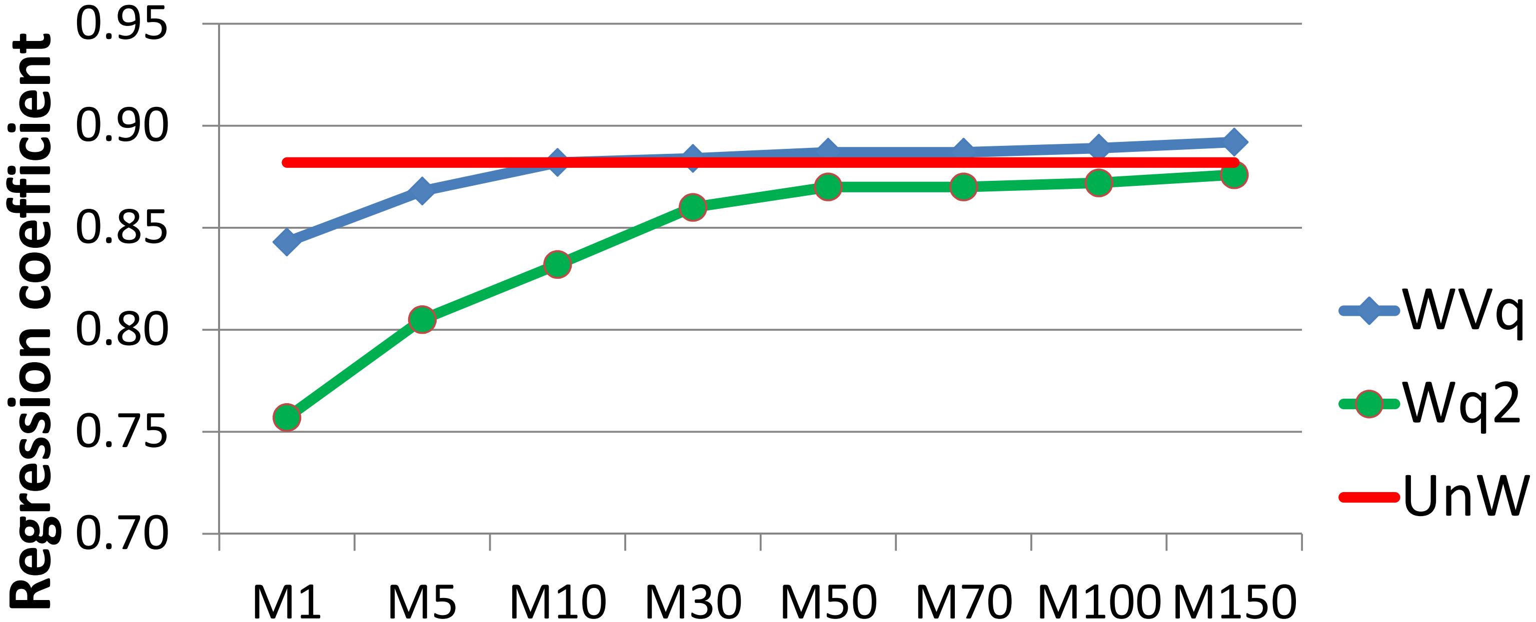
Reliability of GEBV from GBLUP with original (UNW) or weighted (WV_q and Wq^2) G- matrices, averaged over 5 traits



M1=a weight on a single marker; M5 ~ M150= a common weight on 5 ~ 150 markers

WV_q lead to reliability 1.7% higher than original GBLUP

Regression of DRP on GEBV from GBLUP with original (UNW) or weighted (WV_q and Wq^2) G-matrices, averaged over 5 traits



M1=a weight on a single marker, M5 ~ M150= a common weight on 5 ~150 markers

WV_q leads to less bias than Wq^2

Group weighting has less bias than single-marker weighting

Reliability of GEBV from GBLUP with G-matrix weighted by WV_q and a common weight on a group of 30 markers, derived from 0, 1, 3, or 5 year old data

Traits	Lag 0	Lag 1y	Lag 3y	Lag 5y
Milk	51.1	51.1	50.8	50.6
Fat	50.5	50.5	50.8	49.9
Protein	47.2	47.0	46.9	46.7
Fert	44.9	44.8	44.6	44.5
Masti	40.1	40.2	40.3	40.3
Mean	**46.8**	**46.7**	**46.7**	**46.4**

Weight derived from 3y old data does not reduce reliability

Regression of DRP on GEBV using a G-matrix weighted by WVq derived from 0, 1, 3, or 5 year old data

Trait	**Group-marker(30) weight**			
	Lag 0	Lag 1y	Lag 3y	Lag 5y
Milk	0.87	0.87	0.87	0.87
Fat	0.82	0.82	0.82	0.81
Protein	0.82	0.82	0.82	0.81
Fert	0.98	0.98	0.97	0.94
Masti	0.90	0.90	0.88	0.89
Mean	**0.88**	**0.88**	**0.87**	**0.86**

Weights derived from old data did not increase bias

Conclusions

- Posterior variance of individual SNP from BMIX is a good alternative weighting factor.
- A common weight on a group of about 30 markers could be a good weighting method.
- Weights can be updated once per two or three years.

Thank you for your attention